Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 42 (2007) 606–613
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Optimisation of correlation weights of SMILES invariants for modelling oral quail toxicity

Andrey A. Toropov*, Emilio Benfenati

Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche 'Mario Negri', Via Eritrea 62, 20157 Milan, Italy

Received 1 September 2006; accepted 30 November 2006
Available online 15 December 2006

Abstract

The SMILES (simplified molecular input line entry system) nomenclature was used to elucidate the molecular structure in constructing the quantitative structure–property/activity relationships (QSPR/QSAR) for predicting quail toxicity after oral exposure. The presence of chemical elements in different electronic states (e.g., C, c, O, o, Cl, Br, etc.) and of different covalent bonds (i.e., –,=, and #) was used as local invariants. Combinations of different presence/absence codes for local features of the SMILES were used as global invariants. The statistical characteristics of this model are $n = 97$, $r^2 = 0.755$, $s = 0.445$, $F = 293$ (training set); $n = 18$, $r^2 = 0.731$, $s = 0.587$, $F = 43$ (test set).

Keywords: QSAR; Pesticide; Quail oral toxicity; SMILES; Optimal descriptor

1. Introduction

Predicting ecotoxicity is an important issue from the chemical, biochemical, agricultural, and ecological points of view. There are a number of investigations on quantitative structure–property/activity relationships (QSPR/QSAR) modelling of fish [1,2], *Daphnia magna* [3,4], bee [5], mouse and rats [6]. Few articles, on bird toxicity, contain models based on the molecular architecture of substances [7,8]. The limited amount of data on bird toxicity contributes to this situation, making more difficult the task of QSAR modelling for birds.

The aim of the present investigation was to estimate the utility of optimal descriptors calculated with the SMILES nomenclature [9,10] in QSAR analysis of data on quail oral toxicity. Toxicity towards birds is used to assess the impact of the pesticides on terrestrial animals, which may be exposed to the pesticide. In the case of oral toxicity study, the test substance is administered as a single oral dose either by gavage or capsule [11]. Vice versa, in the case of dietary studies, the administration is through the diet. Within the EC funded project DEMETRA [12] we found a relatively low correlation ($r^2 = 0.6$) between the toxicity values reported in the two cases for the pesticides taken into consideration. This difference may be due to the result of several causes of course natural variability, but in addition issues related to the specificity of the test. In the case of dietary study, the animal can avoid food, while for oral studies it is obliged to eat. However, after a while the animal can throw up the administered chemical. Oral study is performed through a single administration of the chemical, while dietary lasts longer. In this case there may be also stability problems of the chemical in the feed.

For all these reasons the two tests were studied separately.

1.1. Toxicity data

As endpoint we used the decimal logarithm $\log(1/C)$, where C is the concentration, in mmol/kg, expressed as LC_{50}-96h, which is the dose that kills 50% of quails in 96 h. The pesticides were administered orally. The data set contains various pesticides which is a complex case. Brian Montague kindly provided the toxicity data for bird quail oral exposure from the US EPA-OPP. Toxicity values were collected and

* Corresponding author. Sergeli 8-A, Home 4, Room 6, 700085 Tashkent, Uzbekistan.
E-mail address: aatoropov@yahoo.com (A.A. Toropov).

0223-5234/$ - see front matter © 2006 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2006.11.018

compared with the BBA database, as described in Ref. [13], in order to use the most reliable data in case of multiple values and to eliminate compounds in case of disagreement between values. Indeed, it is quite common that for the same compound more than one value is reported, in the EPA-OPP database, or by comparing data between BBA and EPA. If the toxicity values varied by a factor of more than 4, we eliminated the compounds. We obtained a data set of 115 pesticides which were sorted on the basis of toxicity values, and split into a training ($n = 96$) and test set ($n = 19$) using one molecule out of six for the test set.

2. Method

The optimal descriptor [14,15] for the model toxicity has been defined as

$$\mathrm{DCW} = \mathrm{DCW}\left(\mathrm{SMILES}_j\right) = \mathrm{CW}\left(\mathrm{GI}_j\right) \cdot \Pi_{k=1}^{N} \mathrm{CW}(\mathrm{LI}_k) \qquad (1)$$

where GI_j is a global SMILES invariant for the jth compound (scheme of constructing the GI_k is shown in Table 1), $\mathrm{CW}(\mathrm{GI}_j)$ is the correlation weight of the GI_j; LI_k is a local SMILES invariant (e.g., C, c, N, n,…, =, #,…, C=C, [N+], etc.) and $\mathrm{CW}(\mathrm{LI}_k)$ is the correlation weight of the LI_k; N is the number of local invariants in the jth SMILES. The SMILES invariants are detected by hierarchy: firstly selection of a four-character fragment ([O−], [Sn]), then, secondly, a three-character fragment (C=C, C#C), thirdly, a two-character fragment (Cl, Br), and finally all remaining characters.

The correlation coefficient between the DCW(SMILES) and $\log(1/C)$ is a mathematical function of the CWs and can be expressed as

$$R\left(\mathrm{DCW}\left(\mathrm{SMILES}_j\right),\ \log\left(1/C_j\right)\right) = F\left(\mathrm{CW}\left(\mathrm{GI}_j\right),\ \left\{\mathrm{CW}(\mathrm{LI}_1), \mathrm{CW}(\mathrm{LI}_2), \ldots, \mathrm{CW}(\mathrm{LI}_N)\right\}\right) \qquad (2)$$

The Monte Carlo method can be used to calculate the numerical values of the $\mathrm{CW}^*(\mathrm{GI}_j)$, $\{\mathrm{CW}^*(\mathrm{LI}_1)$, $\mathrm{CW}^*(\mathrm{LI}_2)$,…, $\mathrm{CW}^*(\mathrm{LI}_N)\}$ which produce as large as possible value of $R(\mathrm{DCW}(\mathrm{SMILES}_j), \log(1/C_j))$ on a training set. With the values of the $\mathrm{CW}^*(\mathrm{GI}_j)$, $\{\mathrm{CW}^*(\mathrm{LI}_1)$, $\mathrm{CW}^*(\mathrm{LI}_2)$,…, $\mathrm{CW}^*(\mathrm{LI}_N)\}$, one can define the toxicity model, by the least squares method, as

$$\mathrm{Log}(1/C) = C_0 + C_1 \cdot \mathrm{DCW} \qquad (3)$$

The predictive ability of this model can be estimated with an external test set, containing compounds never used to build up the model.

Table 2
Statistical characteristics of OCWLGI models for training and test sets

LimNI	Probe no.	Training set, $n = 96$			Test set, $n = 19$			A	B
		r^2	s	F	r^2	s	F		
0	1	0.8560	0.3421	559	0.4691	0.7658	15	67	0
	2	0.8577	0.3401	566	0.4512	0.7813	14		
	3	0.8568	0.3412	562	0.4433	0.7846	14		
1	1	0.8168	0.3859	419	0.4292	0.7939	13	45	22
	2	0.8144	0.3884	412	0.4514	0.7804	14		
	3	0.8159	0.3869	417	0.4374	0.7881	13		
2	1	0.8041	0.3990	386	0.4572	0.7716	14	33	34
	2	0.8024	0.4008	382	0.4512	0.7695	14		
	3	0.8029	0.4003	383	0.4621	0.7642	15		
3	1	0.7879	0.4153	349	0.5002	0.7330	17	30	37
	2	0.7879	0.4152	349	0.4964	0.7350	17		
	3	0.7858	0.4173	345	0.5151	0.7207	18		
4	*1*	*0.7823*	*0.4206*	*338*	*0.6135*	*0.6677*	*27*	*27*	*40*
	2	*0.7824*	*0.4206*	*338*	*0.6132*	*0.6683*	*27*		
	3	*0.7810*	*0.4220*	*335*	*0.6080*	*0.6702*	*26*		
5	1	0.7048	0.4899	224	0.4783	0.7789	16	26	41
	2	0.7038	0.4907	223	0.4914	0.7707	16		
	3	0.7030	0.4913	223	0.4926	0.7707	17		
6	1	0.6922	0.5002	211	0.3060	0.8821	7	25	42
	2	0.6925	0.5000	212	0.3035	0.8830	7		
	3	0.6923	0.5001	212	0.3088	0.8799	8		
7	1	0.5879	0.5788	134	0.3296	0.8675	8	23	44
	2	0.5880	0.5787	134	0.3314	0.8669	8		
	3	0.5880	0.5787	134	0.3282	0.8684	8		
8	1	0.5834	0.5820	132	0.3375	0.8601	9	21	46
	2	0.5829	0.5823	131	0.3296	0.8628	8		
	3	0.5835	0.5819	132	0.3363	0.8601	9		
9	1	0.5467	0.6070	113	0.4549	0.8073	14	19	48
	2	0.5459	0.6076	113	0.4407	0.8110	13		
	3	0.5453	0.6080	113	0.4377	0.8127	13		

The r, s, and F are correlation coefficient, standard error, and Fisher F-ratio, respectively. A and B are the numbers of active and blocked SMILES invariants. Values in italic show the probes with highest statistical results on the test set.

3. Results and discussion

The statistical characteristics of models obtained in three attempts (probes 1, 2, and 3) at optimisation, without any discrimination of SMILES invariants, are shown in Table 2. The statistical characteristics of the models (without discrimination) on all three attempts for the training set are satisfactory but not for the test set.

Clearly, if an invariant is present only once in the training set it cannot provide sound information for the external set. Generally, a rare invariant (appearing in the training set once, twice,…, $N < \mathrm{LimNI}$) should be blocked since it can cause overtraining; this happens, when the model shows good performances on the training set but not on the test set. If an invariant is rare, we prefer to block its CW. When

Table 1
Calculation of global SMILES invariants

Position in GI_j	2	3	4	5	6	7	8	9	10	11	12
SMILES LI_k	O, o	N, n	F	Cl	Br	I	S, s	P, p	(	=	#

The first character is defined as '&'. If in a given SMILES, an LI_k is present, then in the relevant GI_j position 1 will appear, otherwise 0. For instance, If jth SMILES is "ClC=CCCl", then GI_j is &00010000010.

Table 3
SMILES invariants, their correlation weights (CWs) on three probes of the Monte Carlo method optimisation, and numbers of each invariant in the training set

SMILES invariant	CW of probe 1	CW of probe 2	CW of probe 3	Number in the training set
S	1.0000000	1.0000000	1.0000000	2
N	0.6840241	0.6844927	0.6806766	38
C	1.0168545	1.0214992	1.0148354	448
[Sn]	1.0000000	1.0000000	1.0000000	1
[O−]	1.3032790	1.2291084	1.3679146	15
[N+]	1.3266275	1.4001033	1.2702415	15
S	0.7114531	0.7124271	0.7211745	34
P	1.0000000	1.0000000	1.0000000	2
S=P	2.8073151	2.8056893	2.7431688	15
O	1.1060228	1.1051722	1.1014493	127
N	0.8785108	0.8803765	0.8876252	82
O=S	1.0000000	1.0000000	1.0000000	4
O=P	3.1612473	3.1783571	3.0706878	6
N=N	1.0000000	1.0000000	1.0000000	1
N=C	1.5079651	1.5001250	1.4555279	11
N#C	1.0000000	1.0000000	1.0000000	4
F	1.0698412	1.0687115	1.0564752	58
Cl	1.0525382	1.0570273	1.0437316	91
Br	2.2338355	2.2486664	2.1991777	7
C	1.0698245	1.0710391	1.0676313	510
C=O	0.6384422	0.6489776	0.6471021	33
C=C	0.8413260	0.8582418	0.8534222	8
=	0.8609814	0.8641075	0.8519180	15
C#C	1.0000000	1.0000000	1.0000000	2
=O	0.4204617	0.4164500	0.4149140	39
3	1.0000000	1.0000000	1.0000000	4
2	1.3966197	1.3892871	1.3890431	24
1	0.9389577	0.9275551	0.9376545	208
(	1.0202734	1.0211414	1.0263371	782
&111110000111	1.0000000	1.0000000	1.0000000	1
&11110000110	1.0000000	1.0000000	1.0000000	2
&11101000110	1.0000000	1.0000000	1.0000000	1
&11100010110	1.0000000	1.0000000	1.0000000	2
&11011000110	1.0000000	1.0000000	1.0000000	1
&11010011110	1.0000000	1.0000000	1.0000000	1
&11010010110	1.0000000	1.0000000	1.0000000	2
&11010010100	1.0000000	1.0000000	1.0000000	1
&11010000110	1.3475243	1.3618186	1.3344531	11
&11010000111	1.0000000	1.0000000	1.0000000	1
&11010000100	1.0000000	1.0000000	1.0000000	2
&11001000111	1.0000000	1.0000000	1.0000000	1
&11000011110	2.2434220	2.2821193	2.2032739	9
&11000010110	6.9107925	7.0487469	6.7170762	5
&11000010100	1.0000000	1.0000000	1.0000000	1
&11000000110	3.0816398	3.1090549	3.0010010	9
&11000000100	1.0000000	1.0000000	1.0000000	2
&10100010110	1.0000000	1.0000000	1.0000000	1
&10010011110	1.0000000	1.0000000	1.0000000	1
&10010010110	1.0000000	1.0000000	1.0000000	2
&10010001110	1.0000000	1.0000000	1.0000000	2
&10010000110	1.0592000	1.0363379	1.0219681	7
&10010000100	1.0000000	1.0000000	1.0000000	3
&10000011110	1.9362166	1.9490492	1.8749208	8
&10000000111	1.0000000	1.0000000	1.0000000	1
&10000000110	1.0000000	1.0000000	1.0000000	2
&01100000110	1.0000000	1.0000000	1.0000000	1
&01010000110	1.0000000	1.0000000	1.0000000	2
&01010000100	1.0000000	1.0000000	1.0000000	1
&01010000101	1.0000000	1.0000000	1.0000000	1
&01000010100	1.0000000	1.0000000	1.0000000	1
&01000010011	1.0000000	1.0000000	1.0000000	1
&01000000110	1.0000000	1.0000000	1.0000000	3
&01000000100	1.0000000	1.0000000	1.0000000	3
&00010000100	1.0000000	1.0000000	1.0000000	1
&00010000010	1.0000000	1.0000000	1.0000000	1
&00001000000	1.0000000	1.0000000	1.0000000	1
&00000000100	1.0000000	1.0000000	1.0000000	1

These models were calculated with LimNI = 4.

"CW(x) is blocked", CW(x) is always equal to 1.0, i.e., it does not change the descriptor calculated with Eq. (1).

Previous QSAR studies based on the optimisation of correlation weights of local and global graph invariants (OCWLGI) did not take into account [12,13] the number of a given invariant in the training set. We evaluated the best limitation (LimNI) of the number of a given invariant in the training set considering the statistical quality of the model for the test set.

Table 2 shows the results of optimisation with different LimNI (0–9). The statistical quality of the model on the test

Table 4
Statistical characteristics of OCWLGI models for training and test sets after placing 4-aminopyridine in the training set

LimNI	Probe no.	Training set, $n = 97$			Test set, $n = 18$			A	B
		r^2	s	F	r^2	s	F		
0	1	0.8564	0.3413	567	0.5338	0.7161	18	68	0
	2	0.8574	0.3401	571	0.5336	0.7179	18		
	3	0.8566	0.3411	568	0.5376	0.7139	19		
1	1	0.7904	0.4123	358	0.4868	0.7549	15	45	23
	2	0.7906	0.4122	359	0.5121	0.7409	17		
	3	0.7894	0.4133	356	0.5107	0.7427	17		
2	1	0.7763	0.4260	330	0.5251	0.7255	18	33	35
	2	0.7782	0.4242	333	0.5300	0.7205	18		
	3	0.7761	0.4262	329	0.5307	0.7213	18		
3	1	0.7617	0.4397	304	0.6161	0.6559	26	30	38
	2	0.7627	0.4388	305	0.5815	0.6794	22		
	3	0.7615	0.4399	303	0.5975	0.6737	24		
4	*1*	*0.7549*	*0.4459*	*293*	*0.7307*	*0.5868*	*43*	*27*	*41*
	2	*0.7549*	*0.4459*	*293*	*0.7331*	*0.5744*	*44*		
	3	*0.7550*	*0.4458*	*293*	*0.7293*	*0.5858*	*43*		
5	1	0.6806	0.5090	202	0.5623	0.7204	21	26	42
	2	0.6798	0.5096	202	0.5738	0.7143	22		
	3	0.6803	0.5092	202	0.5637	0.7185	21		
6	1	0.6692	0.5180	192	0.3647	0.8300	9	25	43
	2	0.6680	0.5190	191	0.3722	0.8282	9		
	3	0.6697	0.5177	193	0.3663	0.8297	9		
7	1	0.5678	0.5922	125	0.3870	0.8262	10	23	45
	2	0.5675	0.5923	125	0.3884	0.8254	10		
	3	0.5672	0.5925	125	0.3900	0.8259	10		
8	1	0.5627	0.5956	122	0.3985	0.8194	11	21	47
	2	0.5625	0.5958	122	0.3884	0.8242	10		
	3	0.5622	0.5959	122	0.3868	0.8252	10		
9	1	0.5265	0.6198	106	0.4749	0.7932	14	19	49
	2	0.5260	0.6201	105	0.4626	0.7964	14	19	49
	3	0.5272	0.6193	106	0.4803	0.7887	15	19	49

The r, s, and F are correlation coefficient, standard error, and Fisher F-ratio, respectively. A and B are the numbers of active and blocked SMILES invariants. Values in italic show the probes with highest statistical results on the test set.

Table 5
SMILES invariants, their correlation weights (CWs) on three probes of the Monte Carlo method optimisation, and numbers of each invariant in the training set after placing 4-aminopyridine in the training set

SMILES invariant	CW of probe 1	CW of probe 2	CW of probe 3	Number in the training set
s	1.0000000	1.0000000	1.0000000	2
n	0.6817288	0.6834998	0.6785327	39
c	1.0213201	1.0221599	1.0191011	453
[Sn]	1.0000000	1.0000000	1.0000000	1
[O−]	1.3786927	1.1801161	1.3190622	15
[N+]	1.2786259	1.4463949	1.3162764	15
S	0.7156466	0.7084136	0.7102967	34
P	1.0000000	1.0000000	1.0000000	2
S=P	2.8452106	2.9350326	2.8273511	15
O	1.1059642	1.0970368	1.1060669	127
N	0.8827022	0.8854786	0.8850513	83
O=S	1.0000000	1.0000000	1.0000000	4
O=P	3.1127833	3.2136494	3.1384764	6
N=N	1.0000000	1.0000000	1.0000000	1
N=C	1.4756552	1.4449662	1.4849830	11
N#C	1.0000000	1.0000000	1.0000000	4
F	1.0708643	1.0711587	1.0682364	58
Cl	1.0462934	1.0481633	1.0483499	91
Br	2.2791875	2.2753308	2.2966113	7
C	1.0689522	1.0672434	1.0670763	510
C=O	0.6278263	0.6445596	0.6348883	33
C=C	0.8387036	0.8335769	0.8425620	8
=	0.8694662	0.8879379	0.8715332	15
C#C	1.0000000	1.0000000	1.0000000	2
=O	0.4007385	0.4116357	0.4025736	39
3	1.0000000	1.0000000	1.0000000	4
2	1.3917071	1.4009532	1.4030514	24
1	0.9212871	0.9224296	0.9234200	210
(	1.0202330	1.0201048	1.0217884	782
&11110000111	1.0000000	1.0000000	1.0000000	1
&11110000110	1.0000000	1.0000000	1.0000000	2
&11101000110	1.0000000	1.0000000	1.0000000	1
&11100010110	1.0000000	1.0000000	1.0000000	2
&11011000110	1.0000000	1.0000000	1.0000000	1
&11010011110	1.0000000	1.0000000	1.0000000	1
&11010010110	1.0000000	1.0000000	1.0000000	2
&11010010100	1.0000000	1.0000000	1.0000000	1
&11010000110	1.2272290	1.1831653	1.2070024	11
&11010000111	1.0000000	1.0000000	1.0000000	1
&11010000100	1.0000000	1.0000000	1.0000000	2
&11001000111	1.0000000	1.0000000	1.0000000	1
&11000011110	2.2741796	2.2378620	2.3221629	9
&11000010110	7.0560330	7.1222910	7.1294787	5
&11000010100	1.0000000	1.0000000	1.0000000	1
&11000000110	3.1030328	3.0904822	3.1032532	9
&11000000100	1.0000000	1.0000000	1.0000000	2
&10100010110	1.0000000	1.0000000	1.0000000	1
&10010011110	1.0000000	1.0000000	1.0000000	1
&10010010110	1.0000000	1.0000000	1.0000000	2
&10010001110	1.0000000	1.0000000	1.0000000	2
&10010000110	0.9272841	0.9205979	0.9314774	7
&10010000100	1.0000000	1.0000000	1.0000000	3
&10000011110	1.8907275	1.9085158	1.9770637	8
&10000000111	1.0000000	1.0000000	1.0000000	1
&10000000110	1.0000000	1.0000000	1.0000000	2
&01100000110	1.0000000	1.0000000	1.0000000	1
&01010000110	1.0000000	1.0000000	1.0000000	2
&01010000101	1.0000000	1.0000000	1.0000000	1
&01010000100	1.0000000	1.0000000	1.0000000	1
&01000010100	1.0000000	1.0000000	1.0000000	1
&01000010011	1.0000000	1.0000000	1.0000000	1
&01000000110	1.0000000	1.0000000	1.0000000	3
&01000000100	1.0000000	1.0000000	1.0000000	3
&01000000000	1.0000000	1.0000000	1.0000000	1
&00010000100	1.0000000	1.0000000	1.0000000	1
&00010000010	1.0000000	1.0000000	1.0000000	1
&00001000000	1.0000000	1.0000000	1.0000000	1
&00000000100	1.0000000	1.0000000	1.0000000	1

These models were calculated with LimNI = 4.

Table 6
Building the model with 97 pesticides of the original training set (sub-training)

LimNI	Probe no.	r^2	s	F	r^2	s	F	A	B
4-Aminopyridine in sub-test set (sub-training set, $n = 81$, sub-test set, $n = 16$)									
0	1	0.8553	0.3298	467	0.2233	1.0042	4	63	0
	2	0.8544	0.3308	463	0.2242	1.0020	4		
	3	0.8546	0.3306	464	0.1995	1.0199	3		
1	1	0.8132	0.3746	344	0.3859	0.9454	9	41	22
	2	0.8118	0.3760	341	0.3806	0.9466	9		
	3	0.8121	0.3757	341	0.3711	0.9515	8		
2	1	0.7754	0.4108	273	0.4692	0.8182	12	32	31
	2	0.7752	0.4110	272	0.4610	0.8232	12		
	3	0.7745	0.4117	271	0.4536	0.8299	12		
3	*1*	*0.7677*	*0.4178*	*261*	*0.4977*	*0.8278*	*14*	*30*	*33*
	2	*0.7671*	*0.4184*	*260*	*0.5074*	*0.8308*	*14*		
	3	*0.7664*	*0.4189*	*259*	*0.4859*	*0.8411*	*13*		
4	1	0.7327	0.4481	217	0.3826	0.8871	9	26	37
	2	0.7333	0.4476	217	0.3825	0.8868	9		
	3	0.7333	0.4476	217	0.3789	0.8925	9		
5	1	0.7323	0.4485	216	0.3822	0.8942	9	25	38
	2	0.7334	0.4476	217	0.3815	0.8861	9		
	3	0.7333	0.4477	217	0.3788	0.8933	9		
6	1	0.7236	0.4558	207	0.2953	0.9521	6	24	39
	2	0.7239	0.4555	207	0.2728	0.9763	5		
	3	0.7238	0.4556	207	0.2831	0.9643	6		
7	1	0.6019	0.5469	119	0.2836	0.9454	6	21	42
	2	0.6022	0.5467	120	0.2821	0.9466	6		
	3	0.6020	0.5469	119	0.2760	0.9503	5		
8	1	0.5506	0.5811	97	0.3044	1.0101	6	17	46
	2	0.5521	0.5801	97	0.3022	1.0140	6		
	3	0.5548	0.5784	98	0.3032	1.0052	6		
9	1	0.5543	0.5787	98	0.3057	1.0001	6	17	46
	2	0.5553	0.5780	99	0.3080	0.9839	6		
	3	0.5517	0.5804	97	0.3033	1.0190	6		
4-Aminopyridine in sub-training set (sub-training set, $n = 82$, sub-test, set $n = 15$)									
0	1	0.8557	0.3292	475	0.2725	0.9615	5	64	0
	2	0.8567	0.3281	478	0.3234	0.9337	6		
	3	0.8572	0.3275	480	0.2728	0.9669	5		
1	1	0.7853	0.4016	293	0.4295	0.8970	10	41	23
	2	0.7854	0.4015	293	0.4148	0.9059	9		
	3	0.7842	0.4026	291	0.4064	0.9108	9		
2	1	0.7455	0.4372	234	0.5076	0.7766	13	32	32
	2	0.7455	0.4372	234	0.5137	0.7721	14		
	3	0.7443	0.4382	233	0.5178	0.7702	14		
3	*1*	*0.7344*	*0.4466*	*221*	*0.5875*	*0.7676*	*19*	*30*	*34*
	2	*0.7344*	*0.4466*	*221*	*0.5810*	*0.7717*	*18*		
	3	*0.7336*	*0.4473*	*220*	*0.5657*	*0.7850*	*17*		

(*continued on next page*)

Table 6 (*continued*)

LimNI	Probe no.	r^2	s	F	r^2	s	F	A	B
4	1	0.7005	0.4743	187	0.4547	0.8303	11	26	38
	2	0.7000	0.4747	187	0.4500	0.8349	11		
	3	0.7003	0.4744	187	0.4546	0.8292	11		
5	1	0.7005	0.4743	187	0.4498	0.8327	11	25	39
	2	0.7003	0.4745	187	0.4514	0.8319	11		
	3	0.7001	0.4746	187	0.4477	0.8328	11		
6	1	0.6912	0.4816	179	0.3382	0.9081	7	24	40
	2	0.6919	0.4810	180	0.3238	0.9236	6		
	3	0.6909	0.4819	179	0.3345	0.9126	7		
7	1	0.5751	0.5649	108	0.3296	0.9069	6	21	43
	2	0.5752	0.5648	108	0.3140	0.9175	6		
	3	0.5752	0.5649	108	0.3033	0.9276	6		
8	1	0.5334	0.5920	91	0.3666	0.9101	8	17	47
	2	0.5297	0.5944	90	0.3657	0.9224	7		
	3	0.5293	0.5946	90	0.3668	0.9274	8		
9	1	0.5303	0.5939	90	0.3700	0.9213	8	17	47
	2	0.5306	0.5938	90	0.3658	0.9246	7		
	3	0.5291	0.5947	90	0.3640	0.9292	7		

Values in italic show the probes with highest statistical results on the test set.

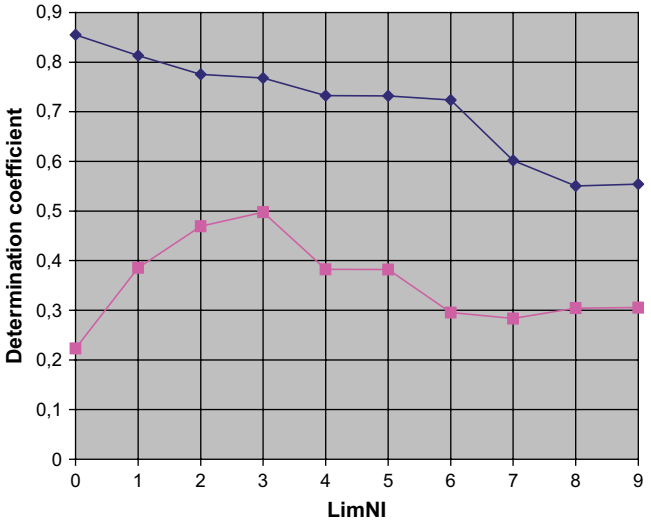

Fig. 2. Determination coefficients of the sub-training and the sub-test sets (4-aminopyridine in the sub-test set) for different LimNI values. Rhombus corresponds to the training set and the square corresponds to the test set.

set is a function of LimNI. The best model (for the test set) was with LimNI = 4. The number of active invariants (A), in the model and the number of blocked SMILES invariants (B) for each LimNI are shown in Table 2. The correlation weights with LimNI = 4 are presented in Table 3. As mentioned, if an invariant occurs for less than four times, it is blocked, so the related CW equals to 1.

4-Aminopyridine is an influential outlier in the training and test sets. Probably, the features of 4-aminopyridine, that cause extraordinary toxicity are geometrical and quantum chemical ones. For instance, a high level of symmetry and an extremely compact arrangement of atoms can cause 'specific' toxicity. If this compound is placed in the training set, the statistical characteristics of the OCWLGI models for the test set improve. The statistical characteristics of OCWLGI models in this case are presented in Table 4. The results obtained for quail are quite good, considering the wide diversity of the chemical compounds studied, which include many different chemical structures and presence of functional groups. The compounds we studied, pesticides, are typically active compounds, and for their diversity they act through a series of different toxic mechanisms. A recent report on QSAR and pesticides [16] reported that some values were obtained in the case of aquatic toxicity of pesticides. r^2 was 0.36 for fish and 0.28 for *Daphnia*, using some simple models. In the case of fish toxicity of a heterogeneous set of compounds (including some pesticides) the r^2 was lower than 0.7 [17].

Correlation weights for this model are presented in Table 5. Fig. 1 gives a diagram of correlations for the training and test sets, for different LimNI.

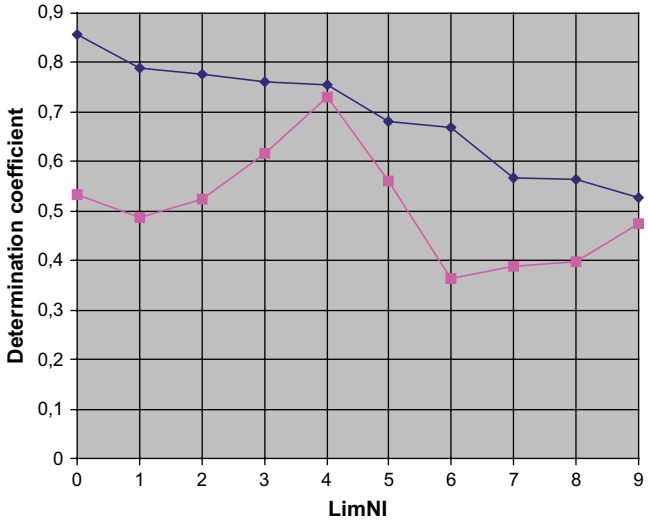

Fig. 1. Determination coefficients of the training and test sets (after placing 4-aminopyridine in the training set) for different LimNI values. Rhombus corresponds to the training set and the square corresponds to the test set.

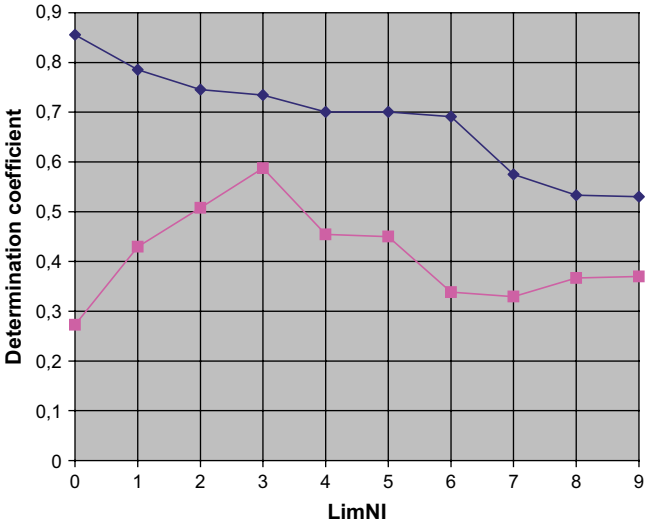

Fig. 3. Determination coefficients of the sub-training and the sub-test sets (4-aminopyridine in the sub-training set) for different LimNI values. Rhombus corresponds to the training set and the square corresponds to the test set.

Table 7
Training and test sets of pesticides after placing 4-aminopyridine in the training set

No	SMILES (training set, $n = 97$)	DCW	Expr.	Calc.
1	Nc1ccncc1	0.56758	0.79764	−0.68627
2	O=C1N(C(=O)C(N1Cl)(C)C)Cl	0.34259	−0.93973	−0.75196
3	ClC=CCCl	0.98146	−0.13664	−0.56541
4	O=C(N(c1c(cccc1CC)CC)COC)CCl	1.36345	−0.74476	−0.45387
5	O=C(ON=CC(SC)(C)C)NC	7.18831	1.97841	1.24699
6	CSc1nc(nc(n1)NC(C)C)NCC	0.26877	−0.99543	−0.77352
7	CN(C=Nc1ccc(cc1C)C)C=Nc1c(cc(cc1)C)C	2.92209	−0.42899	0.00125
8	Clc1nc(nc(n1)NC(C)C)NCC	0.36760	−0.63924	−0.74466
9	O=C(Oc1cccc2c1OC(O2)(C)C)NC	6.39072	1.07004	1.01409
10	O=P(SCCS(=O)CC)(OC)OC	2.48579	−0.96068	−0.12615
11	S=P(OC(C(Cl)(Cl)Cl)Cl)(OCC)OCC	8.40804	1.07919	1.60315
12	N=C(NC(=N)NCCCCCCNC(=N)NC(=N)Nc1ccc(cc1)Cl)Nc1ccc(cc1)Cl	0.74288	−0.50758	−0.63508
13	S=P(Oc1nc(c(cc1Cl)Cl)Cl)(OCC)OCC	4.34447	1.03966	0.41659
14	C#CCOC(=O)C(Oc1ccc(cc1)Oc1ncc(cc1F)Cl)C	0.57660	−0.61909	−0.68363
15	O=C1N(OCC1(C)C)Cc1ccccc1Cl	0.97300	−1.01997	−0.56788
16	O[Sn](C1CCCCC1)(C1CCCCC1)C1CCCCC1	2.43309	0.10860	−0.14154
17	N#Cc1c(cccc1Cl)Cl	1.09766	−0.59887	−0.53148
18	O=C(C(Oc1ccc(cc1Cl)Cl)C)O	0.96764	−0.17781	−0.56945
19	O=P(OC=C(Cl)Cl)(OC)OC	4.98196	1.39987	0.60273
20	[O−][N+](=O)c1cc(c(c(c1)Cl)N)Cl	0.94727	−0.63823	−0.57540
21	Clc1(c(c(c(c1Cl)Cl)Cl)Cl)c1(c(c(c(c1Cl)Cl)Cl)Cl)Cl	1.92719	−0.17186	−0.28926
22	O=C(N(c1c(scc1C)C)C(COC)C)CCl	1.10864	−0.58793	−0.52828
23	N=C(NCCCCCCCCCCCC)N	2.66386	−0.38030	−0.07415
24	ClC=CCN12CN3CN(C1)CN(C2)C3	1.51325	−0.75841	−0.41013
25	S=P(SCSP(=S)(OCC)OCC)(OCC)OCC	5.48625	0.47840	0.74999
26	O=C1C(=NN(c2ccc(cc2)Cl)C(=C1)C)C(=O)O	0.66957	−1.17287	−0.65649
27	S=P(Oc1cc(c(cc1)SC)C)(OC)OC	7.68873	1.59335	1.39311
28	O=C(Oc1cc(ccc1)N=CN(C)C)NC	3.28710	0.79310	0.10783
29	FC(c1ccc(cc1)C=CC(=NN=C1NCC(CN1)(C)C)C=Cc1ccc(cc1)C(F)(F)F)	1.80248	−0.56778	−0.32568
30	O=C1N(C(=O)NC(C)C)CC(=O)N1c1cc(cc(c1)Cl)Cl	0.14452	−0.44972	−0.80980
31	S=P(Oc1c(cccc1)C(=O)OC(C)C)(NC(C)C)OCC	6.91219	1.59885	1.16636
32	O=S(=O)(C(C(C(C(C(C(C(C(F)(F)F)(F)F)(F)F)(F)F)(F)F)(F)F)(F)F)(F)F)(F)	4.97658	1.08095	0.60116
33	O=C(C(Oc1c(cc(cc1)Cl)C)C)O	1.02901	−0.51767	−0.55153
34	O=C(ON=C(SC)C)NC	6.04382	0.82634	0.91280
35	[O−][N+](=O)N=C1N(Cc2ccc(nc2)Cl)CCN1	1.78913	0.22501	−0.32957
36	O=C1N(SC=C1)CCCCCCCC	3.53489	−0.25518	0.18019
37	O=S(=O)(c1cc(c(c(c1)[N+](=O)[O−])N(CCC)CCC)[N+](=O)[O−])N	2.17718	−0.16516	−0.21626
38	[O−][N+](=O)c1ccc(cc1)O	2.53022	−0.61779	−0.11318
39	S=P(Oc1ccc(cc1)[N+](=O)[O−])(OC)OC	7.98927	1.54181	1.48087
40	Clc1c(c(c(c(c1O)Cl)Cl)Cl)Cl	1.56807	−0.37186	−0.39412
41	S=P(SCSCC)(OCC)OCC	5.82275	1.57058	0.84824
42	O=C(Nc1cc(c(cc1)Cl)Cl)CC	0.92419	0.03544	−0.58214
43	S=P(Oc1cnc(nc1)C(C)(C)C)(OC(C)C)OCC	8.70643	1.19550	1.69028
44	S=P(Oc1ccc(cc1)Sc1ccc(cc1)OP(=S)(OC)OC)(OC)OC	6.75442	1.23111	1.12029
45	S=P(SCSC(C)(C)C)(OCC)OCC	7.20846	1.00375	1.25287
46	O=C(c1c(nc(c(c1CC(C)C)C1=NCCS1)C(F)(F)F)C(F)F)OC	0.78667	−0.68356	−0.62229
47	O=P(SCCCC)(SCCCC)SCCCC	5.20200	0.31874	0.66698
48	Clc1c(ccc(c1)Cl)Oc1ccc(cc1O)Cl	1.46609	−0.45475	−0.42390
49	O=C(Oc1cc(c(c(c1)C)C)C)NC	2.80783	−0.09041	−0.03211
50	OC(C(=Cc1ccc(cc1)Cl)n1ncnc1)C(C)(C)C	0.65011	−0.69955	−0.66217
51	OCn1nc(cc1C)C	0.59090	−0.77925	−0.67946
52	O=C(OC(C)C)COc1c(cc(cc1)Cl)Cl	1.15088	−0.85376	−0.51594
53	O=C1C(=C(SS1)Cl)Cl	0.32159	−0.12069	−0.75810
54	O=C(NNc1cc(ccc1OC)c1ccccc1)OC(C)C	2.53681	−0.53599	−0.11125
55	CN(CCCCCCCCCC)(CCCCCCCCCC)C	4.14662	0.22234	0.35881
56	O=C(c1cnc(cc1)C(=O)OCCC)OCCC	1.10429	−0.73012	−0.52955
57	ClC(c1nsc(n1)OCC)(Cl)Cl	0.71798	−0.35454	−0.64235
58	[O−][N+](=O)c1c(c(cc(c1Nc1ncc(cc1Cl)C(F)(F)F)[N+](=O)[O−])C	0.76833	−0.58335	−0.62765
59	c1c2c(ccc1)cccc2	2.11302	−1.32193	−0.23500
60	Clc1ccc(cc1C)O	1.24026	−1.03343	−0.48984
61	S=P(Oc1nc(nc(c1)C)N(CC)CC)(OC)OC	6.46158	0.88278	1.03478
62	C#CCC1=C(C(OC(=O)C2C(C2C=C(C)C)(C)C)CC1=O)C	0.73617	−0.59081	−0.63704
63	O=C(N(SN(C(=O)ON=C(SC)C)C)C)ON=C(SC)C	2.73866	−0.75634	−0.05231
64	O=C1N(C(=O)NC(=O)N1Cl)Cl	0.10189	−0.90467	−0.82225

(*continued on next page*)

Table 7 (*continued*)

No	SMILES (training set, $n = 97$)	DCW	Expr.	Calc.
65	S=P(SCN1N=Nc2c(cccc2)C1=O)(OC)OC	5.50404	0.98304	0.75518
66	[O−][N+](=O)c1cc(cc(c1N(c1c(cc(cc1Br)Br)Br)C)C(F)(F)F)[N+](	9.73934	2.09912	1.99189
67	O=C1N(C(=O)C(N1Br)(C)C)Cl	0.60811	−0.64650	−0.67443
68	N#Cc1cc(c(c(c1)Br)OC(=O)CCCCCC)Br	4.15170	0.03497	0.36030
69	OC(c1ccc(cc1)Cl)(Cn1ncnc1)C(C1CC1)C	0.49682	0.28901	−0.70693
70	S=P(Oc1nc(nc(c1)C)C(C)C)(OCC)OCC	7.82498	1.76743	1.43289
71	O=C(c1c(ccc(c1OC)Cl)Cl)O	0.90523	0.01002	−0.58767
72	O=C(C(Oc1ccc(cc1)Oc1c(cc(cc1)Cl)Cl)C)OC	1.19392	−1.11044	−0.50338
73	O=S1OCC2C(C3(C(C2(C(=C3Cl)Cl)Cl)(Cl)Cl)Cl)CO1	5.31567	0.98626	0.70018
74	O=C(C1C(C2OC1CC2)C(=O)O)O	1.00627	−0.42379	−0.55817
75	O=P(NC(C)C)(Oc1cc(c(cc1)SC)C)OCC	10.26672	2.27790	2.14588
76	S=P(Oc1cc(c(cc1)[N+](=O)[O−])C)(OC)OC	8.88923	1.06998	1.74366
77	[O−][N+](=O)c1c(c(cc(c1)C(F)(F)F)[N+](=O)[O−])N(CCCl)CCC	1.12077	−1.29398	−0.52474
78	O=C1C=C(ON1)C	1.50639	−1.17389	−0.41213
79	BrC	2.43634	0.11413	−0.14059
80	N1=CNc2c(ncnc12)NCc1ccccc1	0.56955	−0.85113	−0.68569
81	Clc1c(ccc(c1)Cl)C1(Cn2ncnc2)OC(CO1)CCC	1.31189	−0.91667	−0.46893
82	O=S(=O)(C(C(C(C(C(C(C(F)(F)F)(F)F)(F)F)(F)F)(F)F)(F)F)(F)F)(F)	4.53859	0.04623	0.47327
83	N#CSCSC1=Nc2ccccc2S1	0.59970	−0.44283	−0.67689
84	O=C(C(c1ccc(cc1)Cl)(c1ccc(cc1)Cl)O)OCC	1.07952	−0.27104	−0.53678
85	Nc1nc(nc(n1)N)NC1CC1	0.22132	−1.03095	−0.78737
86	CN(CCCCCCCCCC)(CCCCCCC(C)C)C	4.03770	1.01016	0.32701
87	O=C(OC1OC(OC(C1)C)C)C	1.29645	−0.95893	−0.47344
88	[O−][N+](=O)c1cc(cc(c1O)C(CC)C)[N+](=O)[O−]	2.63184	0.77859	−0.08350
89	N#CC(c1cc(ccc1)Oc1ccccc1)OC(=O)C(c1ccc(cc1)Cl)C(C)C	0.86981	0.04225	−0.59802
90	OCCN1CN(CN(C1)CCO)CCO	1.55911	−0.84075	−0.39674
91	O=C(C(Oc1ccc(cc1C)Cl)C)O	0.98860	−0.40544	−0.56333
92	O=C(N(c1c(cccc1C)C)C(C(=O)OC)C)COC	1.69124	−0.54549	−0.35816
93	O=P(SC)(N)OC	6.12268	1.14536	0.93582
94	O=C(c1c(cccc1Cl)Sc1nc(cc(n1)OC)OC)O	0.35265	−0.66135	−0.74903
95	S=P(Oc1ccc(cc1)SC)(SCCC)OCC	5.46173	0.83642	0.74283
96	O=P(C(C(Cl)(Cl)Cl)O)(OC)OC	7.69413	1.06043	1.39469
97	O=C1N(C(=O)N(C(=O)N1Cl)Cl)Cl	0.11097	−0.85750	−0.81960
No	SMILES (test set, $n = 18$)	DCW	Expr.	Calc.
1	S=P(OC(C)C)(OC(C)C)SCCNS(=O)(=O)c1ccccc1	1.19998	−0.54235	−0.50161
2	O=C(Oc1cccc2c1OC(C2)(C)C)NC	6.17685	1.64251	0.95164
3	O=C1Sc2c(nc3c(n2)ccc(c3)C)S1	2.47375	0.07755	−0.12966
4	O=C1C(C(=O)c2c1cccc2)C(=O)C(c1ccc(cc1)Cl)c1ccccc1	0.27017	−0.12077	−0.77311
5	N#CC(C(=O)N)(Br)Br	2.36782	−0.16542	−0.16060
6	O=C(c1cc(ccc1)C)N(CC)CC	2.60729	−0.85659	−0.09067
7	Clc1ccc(cc1)Cl	1.09766	−1.03897	−0.53148
8	O=C(N(c1ccccc1)C(C)C)CCl	1.00931	0.38126	−0.55728
9	O=C(NCCCC)OCC#CI (?)	0.89040	−0.42561	−0.59200
10	O=S1(=O)N(C(=O)c2c(cccc2)N1)C(C)C	2.62782	−0.64982	−0.08468
11	OCNCCO (?)	1.31878	−1.28164	−0.46692
12	S=P(Oc1cc2c(cc1)C(=C(C(=O)O2)Cl)C)(OCC)OCC	6.27714	2.18674	0.98092
13	O=C1N(C(=NC(=O)N1C1CCCCC1)N(C)C)C	0.72563	−0.95035	−0.64012
14	O=C(Oc1cc(c(c(c1)C)SC)C)NC	4.56923	1.06058	0.48222
15	O=C(C(Oc1cccc(c1)Cl)C)O	0.92483	−0.77172	−0.58195
16	S=P(SCCSCC)(OCC)OCC	6.22424	1.35928	0.96548
17	O=C(COc1c(cc(cc1)Cl)C)O	0.92483	−0.27395	−0.58195
18	O=S(=O)(c1ccc(cc1)CCCCCCCCCCCC)O (?)	1.07166	−0.59009	−0.53908

DCW values calculated with CW from Table 6 (LimNI = 4, probe 1), experimental and calculated with Eq. (4) $\log(1/C)$ values. SMILES, which has invariant absent in the training set, are marked by '?'.

We tried to estimate the possibility of defining the LimNI for external substances. We separated compounds of the training set into sub-training and sub-test sets and examined two situations. First, 4-aminopyridine was placed in the sub-test set, and second, the compound was placed in the sub-training set. Results strongly depend on the position of 4-aminopyridine and the best models for the sub-test set are achieved with LimNi = 3 (see Table 6, Figs. 1–3).

Besides 4-aminopyridine, three compounds of the test set contain SMILES invariants that are absent in the training set. Correlation weights of these SMILES invariants were, also blocked, so they equal 1.0. These compounds are marked by '?' in Table 7. Predicted toxicity values for these compounds are satisfactory.

Statistical characteristics of the models (after placing 4-aminopyridine in the training set) with different LimNI are

Table 8
Calculation of the DCW for 4-aminopyridine with CWs of the first OCWLGI probe, LimNI = 4 (Table 6)

No.	SMILES invariant	Correlation weight
1	N	0.88270
2	c	1.02132
3	1	0.92129
4	c	1.02132
5	c	1.02132
6	n	0.68173
7	c	1.02132
8	c	1.02132
9	1	0.92129
10	&01000000000	1.00000
	DCW(Nc1ccncc1)=	*0.56758*

listed in Table 4. These characteristics are more or less similar for all probes of the Monte Carlo method optimisation on each LimNI value. Table 4 also shows that the correlation coefficients between toxicity and the optimal SMILES-based descriptors are related to the LimNI for both training and test sets. Fig. 1 gives a general view of the influence of LimNI. Comparing Table 4 and Fig. 1, one can conclude that the best statistical characteristics of the models on the test set are for LimNI = 4. In other words, the difference between LimNI values in the original training and test sets, and also for the sub-training and sub-test sets, is not large.

The correlation weights obtained by placing 4-aminopyridine in the training set with LimNI = 4 are listed in Table 5. Table 7 shows experimental and calculated toxicity values with correlation weights from Table 5 for the formula.

$$\mathrm{Log}(1/C) = -0.852 + 0.292\mathrm{DCW} \quad (4)$$

The calculation of the DCW for 4-aminopyridine is given in Table 8.

Even blocking a considerable part of the global SMILES invariants and their use in this approach gives a reasonable improvement of the models for both the training and test sets. The list of the global SMILES invariants active in these models provides useful information when searching a mechanistic interpretation of the toxic action. Furthermore, it is probable that, for the given pesticides and this endpoint, local invariants with correlation weights > 1 are promoters of increased toxicity and invariants with correlation weights < 1 are promoters of decreased toxicity. This too can be useful in the search for a mechanistic interpretation of the toxicity. For instance, Tables 2 and 4 both show that the presence of groups such as P=S and P=O is related to toxicity. This is reasonable considering the high toxicity of organophosphates and their sulfurated analogs. The presence of chlorine and bromine is also related to greater toxicity. Since, frequent SMILES invariants increase the model reliability, they can be used to characterize the applicability domain of the model.

It is to be noted that there are publications on using the SMILES in QSPR/QSAR analyses [18,19].

Toxicity endpoints phenomena are effected by numerous factors. Partially these factors are defined by molecular structure and reactions of living organisms. Under such circumstances, high statistical quality of toxicity models is possible only for QSAR of series of compounds with similar structure (limited number of functional groups). Taking into account these circumstances and the variety of examined pesticides (many different functional groups), the statistical characteristics of our model of toxicity towards oral quail are good.

4. Conclusions

SMILES invariants can be useful for developing QSAR models for pesticide toxicity in different cases. Good results were obtained for toxicity of pesticides towards the quail (oral exposure). If a SMILES invariant is poorly represented in the training set it is better not to use it in the model. The occurrence of known SMILES invariants can be used to characterize the applicability domain of the model. In the specific case of quail toxicity, in view of the relatively small number of pesticides, a specific compound may have substantial effect on the results, simply depending upon whether this compound is in the training or the test set.

Acknowledgements

This work was funded by EU contract QLK5-CT-2002-00691, Demetra.

References

[1] C.L. Russom, S.P. Bradbury, S.J. Broderius, D.E. Hammermeister, A. Drummond, Environ. Toxicol. Chem. 16 (1997) 948–967.
[2] H.J.M. Verhaar, C.J. Van Leeuwen, J.L.M. Hermens, Chemosphere 25 (1992) 471–491.
[3] Y. Kamaya, Y. Fukaya, K. Suzuki, Chemosphere 59 (2005) 255–261.
[4] X. Liu, B. Wang, Z. Huang, S. Han, L. Wang, Chemosphere 50 (2003) 403–408.
[5] M. Vighi, M. Masoero, G. Calamari, D. Calamari, Sci. Total Environ. 109-110 (1991) 605–622.
[6] W. Guilian, B. Naibin, Chemosphere 36 (1998) 1475–1483.
[7] J.W. Tas, F. Balk, R.A. Ford, E.J. Van de Plassche, Chemosphere 35 (1997) 2973–3002.
[8] S. Winter, B. Streit, Chemosphere 24 (1992) 1765–1774.
[9] http://www.daylight.com
[10] http://www.epa.gov
[11] http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/850_Ecological_Effects_Test_Guidelines/Drafts/850–2100.pdf
[12] http://www.demetra-tox.net
[13] A. Roncaglioni, E. Benfenati, E. Borani, M.J. Clook, J. Environ. Sci. Health B 39 (2004) 641–652.
[14] A.A. Toropov, E. Benfenati, J. Mol. Struct. (Theochem) 676 (2004) 165–169.
[15] A.A. Toropov, E. Benfenati, J. Mol. Struct. (Theochem) 679 (2004) 225–228.
[16] http://www.mst.dk/homepage/default.asp?Sub=http://www.mst.dk/udgiv/publications/2004/87-7614-434-8/html/default_eng.htm
[17] T.I. Netzeva, A.O. Aptula, E. Benfenati, M.T.D. Cronin, G. Gini, I. Lessigiarska, U. Maran, M. Vracko, G. Schüürmann, J. Chem. Inf. Model. 45 (2005) 106–114.
[18] D. Vidal, M. Thormann, M. Pons, J. Chem. Inf. Model. 45 (2005) 386.
[19] A.A. Toropov, A.P. Toropova, D.V. Mukhamedzhanova, I. Gutman, Indian J. Chem, Sect A. 44 (2005) 1545.